



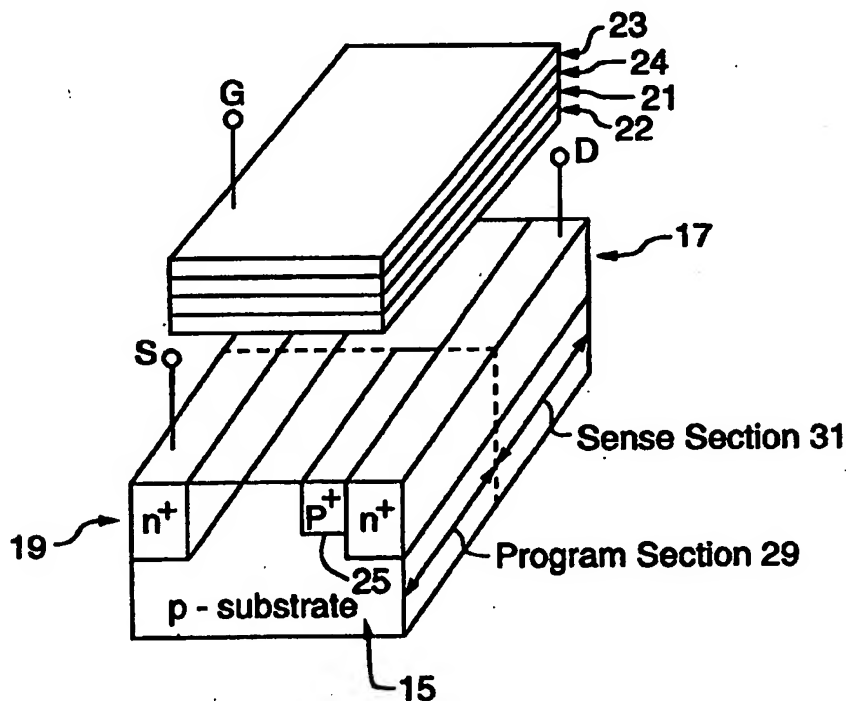
## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>H01L 29/788, 29/10</b>	<b>A1</b>	(11) International Publication Number: <b>WO 97/02605</b> (43) International Publication Date: 23 January 1997 (23.01.97)
(21) International Application Number: PCT/CA96/00446 (22) International Filing Date: 3 July 1996 (03.07.96) (30) Priority Data: 60/001,046 3 July 1995 (03.07.95) US (71)(72) Applicants and Inventors: RANAWEERA, Jeewika, Chandanie [LK/CA]; 20 Vero Crescent, Toronto, Ontario M6M 1S4 (CA). KALASTIRSKY, Ivan [BG/CA]; 151 Springhurst Avenue, Toronto, Ontario M6K 1B9 (CA). GULERSON, Elvira [CA/CA]; 31 Savoy Crescent, Thornhill, Ontario L4J 7W3 (CA). NG, Wai, Tung [CA/CA]; 23 Lincombe Drive, Thornhill, Ontario L3T 3V6 (CA). SALAMA, Clement, Andre, T. [CA/CA]; 66 Castlewood Road, Toronto, Ontario M5N 2L2 (CA). (74) Agent: PERRY, Stephen, J.; Sim & McBurney, 6th floor, 330 University Avenue, Toronto, Ontario M5G 1R7 (CA).		(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  Published With international search report.

(54) Title: METHOD OF FABRICATING A FAST PROGRAMMING FLASH E<sup>2</sup>PROM CELL

## (57) Abstract

In a flash E<sup>2</sup>PROM cell having source and drain regions disposed in a substrate, a channel region intermediate to the source and drain regions, a tunnel dielectric layer overlying the channel region, a floating gate overlying the tunnel dielectric layer, an inter-poly dielectric layer overlying the floating gate and a control gate overlying the inter-poly dielectric layer, the improvement comprising a highly doped p<sup>+</sup> pocket implant covering a portion of the cell width and adjacent to least one of the source and drain regions. The flash E<sup>2</sup>PROM cell is comprised of two sections butted together. The portion (width-wise) covered by the highly doped p<sup>+</sup> pocket implant is referred to as the program section and the remaining portion (width-wise) not covered by the highly doped p<sup>+</sup> pocket implant resembles a conventional E<sup>2</sup>PROM cell and is referred to as the sense section. The highly doped p<sup>+</sup> pocket implanted and the n<sup>+</sup> drain and/or source regions create a junction having narrow depletion width such that in the event the junction is reverse biased, an electric field is created for generating hot electrons for storage on the floating gate, thereby programming the flash E<sup>2</sup>PROM cell when a high positive potential is applied to the control gate. The cell according to the present invention provides short programming time and low operating voltages as compared to prior art devices.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LJ	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

METHOD OF FABRICATING A FAST PROGRAMMING FLASH E<sup>2</sup>PROM CELLField of the Invention

5           The present invention relates in general to  
Electrically Erasable and Programmable Read-Only Memory  
(E<sup>2</sup>PROM), and more particularly to a method of fabricating  
a flash E<sup>2</sup>PROM with fast programming speed and low  
operating voltages.

10

Background of the Invention

Flash memory is a modified form of E<sup>2</sup>PROM which can  
be erased one block at a time and can be programmed one  
15 bit at a time. At the time of filing this application,  
flash memory chips are available in densities of up to  
32M bits. Large data storage capacity with complete  
nonvolatility, results in numerous applications for such  
chips, ranging from cellular telephones, solid-state  
20 disks and memory cards. Such applications have  
traditionally been the domain of ROM, conventional E<sup>2</sup>PROM,  
battery-backed RAM, static RAM (SRAM) and magnetic  
storage.

25           A flash E<sup>2</sup>PROM cell resembles an ordinary MOS (Metal-  
Oxide-Semiconductor) transistor, except for the addition  
of a floating gate, which is buried in the insulator  
between the substrate and the conventional control gate.  
Charge stored on the floating gate alters the threshold  
30 voltage ( $V_{th}$ ) of the device as measured at the control  
gate. Since the floating gate and the control gate are  
both stacked directly above the transistor channel, very  
high densities can be achieved. Even higher densities  
can be obtained by self aligning the control and floating  
35 gates to the source and drain regions.

The cell is programmed and erased by adding  
electrons to and removing electrons from the floating  
gate, respectively. A cell with a high threshold voltage

( $V_{th}$ ) is in its "0" state. After erasure has been completed, the threshold voltage is reduced resulting in a "1" state. The conductivity of the channel determines the information stored in the memory cell (i.e. current flowing through the channel is detected by sense-amplifier circuitry as a "1", while the absence of current is detected as a "0").

Flash E<sup>2</sup>PROM technology has received industry-wide attention recently. Due to its simple single transistor cell architecture, flash memory may eventually cost less to make than DRAM (Dynamic Random Access Memory).

The channel hot electron programming method, used in conventional flash E<sup>2</sup>PROM cells, requires biasing the device at high drain voltage (6 to 8V) to generate hot electrons. This results in additional circuit complexity and cost (ie. an additional external voltage supply), which is a particular disadvantage in mobile applications such as cellular telephones, etc.

Furthermore, the channel hot electron injection method used for programming conventional flash E<sup>2</sup>PROM cells, generates very high lateral source-to-drain currents (in the range of milliamperes). These high currents limit the number of cells that can be programmed at one time.

Existing flash E<sup>2</sup>PROM cells also suffer from slow programming speed ( $\sim 10\mu s$ ) which prevents their widespread application as a replacement for RAMs and electronic hard disks. If the device gate length is scaled down in order to reduce the programming time, then punchthrough between the source and drain is likely to occur.

35

#### Discussion of Prior Art

A well known early prior art predecessor to modern day flash memory devices is the FAMOS (floating-gate avalanche-injection MOS) memory. This device is described in a number of publications, including: Dov Frohman-Bentchowsky, "A Fully Decoded 2048-Bit Electrically Programmable FAMOS Read-Only Memory"; IEEE Journal of Solid State Circuits, Vol. SC-6, No. 5, October 1971, pp 301-306; Ron D. Katznelson and Dov Frohman-Bentchowsky, "An Erase Model for FAMOS EPROM Devices"; IEEE Transactions on Electron Devices, Vol. ED-27, No. 9, September 1980, pp 1744-1752; and U.S. Patent No. 4,203,158 (Frohman-Bentchkowsky, et al).

Another prior art flash memory device is described in Masatada Horiuchi and Hisao Katto, "FCAT-A Low-Voltage High Speed Alternative n-Channel Nonvolatile Memory Device"; Transactions on Electron Devices, Vol. ED-26, No. 6, June 1979, pp. 914-918. The FCAT cell has the p<sup>+</sup> regions placed outside the channel area corners adjacent to the diffused n<sup>+</sup> source and/or drain regions. This method of positioning the p<sup>+</sup> regions results in a corner point injection for the hot electrons which contributes to acute stress in the tunnel oxide. Also, the FCAT cell uses hot hole injection into the thin gate oxide for erasure, which can lead to poor reliability and endurance characteristics.

U.S. Patent 5,464,785 entitled "Method of Making A Flash EPROM Device Having A Drain Edge P<sup>+</sup> Implant", (Hong et al.), describes a flash EPROM structure with a drain edge p<sup>+</sup> implant that utilizes the electric field at the junction between the drain and implant to enhance the generation and injection of hot channel electrons during programming operation. According to Hong et al., the drain edge p<sup>+</sup> implant extends across the entire width of the EPROM cell, thereby preventing lateral current flow through the MOS channel between the source and drain.

Therefore, the data stored in the cell cannot be sensed readily.

The fabrication steps suggested by Hong et. al. are impractical since it is very difficult to etch the narrow vertical trenches (reported typically to be 1000Å wide by 5000Å high) used for the drain edge p<sup>+</sup> implant. Also, the fabrication steps imply a dedicated process which is not CMOS compatible.

No read, write, and erase schemes are described in the Hong et al. patent, nor is any evidence of experimental performance indicated.

#### Summary of the Invention

According to the present invention, a low voltage flash E<sup>2</sup>PROM cell is provided which overcomes the above-identified programming speed limitations of conventional prior art flash memory cells, and in addition allows the cell to operate at low voltages.

More particularly, the flash E<sup>2</sup>PROM cell is comprised of two sections which are butted together. One portion (width-wise) is covered by a highly doped p<sup>+</sup> pocket implant adjacent to at least one of the n<sup>+</sup> source and drain regions. This portion is referred to as the program section. The remaining portion (width-wise) is not covered by the highly doped p<sup>+</sup> pocket implant and resembles a conventional E<sup>2</sup>PROM cell. This portion is referred to as the sense section.

The highly doped p<sup>+</sup> pocket implant and the n<sup>+</sup> drain and/or source regions create a junction having narrow depletion width such that when the junction is reverse biased an electric field is created for generating hot electrons for storage on the floating gate, thereby

programming the flash E<sup>2</sup>PROM. The programming speed of the improved flash E<sup>2</sup>PROM cell of the present invention is in general an order of magnitude faster than a conventional flash E<sup>2</sup>PROM cell of the same dimension. The doping concentration of the p<sup>+</sup> pocket implant is chosen such that the junction will breakdown at low voltages (typically in the range of 2.5 to 5V), thus enabling the drain and/or source regions to be driven directly by the logic level supply voltages. This eliminates the need for large charge pump circuits with high current capacity as used in conventional flash E<sup>2</sup>PROMs. The enhancement of hot electron generation results in smaller drain and/or source currents (typically in the range of 50 to 200  $\mu\text{A}/\mu\text{m}$  of cell width) than prior art devices during the programming operation. Thus, the flash E<sup>2</sup>PROM cell of the present invention consumes much less power, and is more suitable for battery operated portable equipment than similar prior art devices.

High positive and negative voltages are needed only to drive the control gate. Consequently, they can be generated on-chip using simple charge pump circuitry with small current driving capability. The voltages required at the source and drain in all modes of operation can be obtained directly from a low power supply rail.

The flash E<sup>2</sup>PROM cell of the present invention is compatible with CMOS and BiCMOS processes. More particularly, according to one aspect of the present invention, a method of fabricating an E<sup>2</sup>PROM cell is provided in which the fabrication steps are modular and compatible with conventional CMOS and BiCMOS process flows. Cell fabrication can be integrated into existing CMOS and BiCMOS processes with minimal changes. The resulting flash E<sup>2</sup>PROM cell is a self aligned structure. The p<sup>+</sup> pocket implant is carried out in the same way as the n-LDD implant commonly used in modern CMOS and BiCMOS

processes. The generation of hot electrons is dependent on the doping level of the  $p^+/n^+$  junction rather than the size of the channel length. The electric field at this junction ( $\sim 10^6$  V/cm) significantly exceeds the maximum lateral electric field ( $\sim 10^5$  V/cm) available in conventional flash  $E^2$ PROM cells with short channel lengths (0.3 to 0.5  $\mu\text{m}$ ), thus ensuring very efficient injection of hot electrons into the floating gate.

#### 10 Brief Introduction to the Drawings

Detailed descriptions of one prior art flash memory cell and of flash  $E^2$ PROM cells according to a first embodiment and a preferred embodiment of the present invention are provided herein below, with reference to the following drawings, in which:

Figure 1 is a schematic representation of a prior art flash  $E^2$ PROM cell;

Figure 2 is a schematic representation of hot electron injection programming of the prior art flash memory cell depicted in Figure 1;

Figure 3 is a schematic representation of erasure of the prior art flash memory cell depicted in Figure 1 using Fowler-Nordheim tunnelling across the thin gate oxide;

Figure 4 is a schematic representation of the read operation of the prior art flash memory cell depicted in Figure 1;

Figure 5 is a three-dimensional schematic representation of a flash  $E^2$ PROM cell according to a first embodiment of the present invention;



Figure 6 is a three-dimensional schematic representation of Zener and/or avalanche breakdown-based electron injection programming of the flash E<sup>2</sup>PROM cell of Figure 5;

5

Figure 7 is a three-dimensional schematic representation of the read operation of the flash E<sup>2</sup>PROM cell of Figure 5;

10

Figure 8 is a three-dimensional schematic representation of the erasure operation of the flash E<sup>2</sup>PROM cell of Figure 5;

15

Figure 9 is a three-dimensional schematic representation of the preferred embodiment of flash E<sup>2</sup>PROM cell according to the present invention;

20

Figure 10 is a plan view of the device shown in Figure 9;

Figures 11(a) through 11(r) depict the process flow steps for fabricating the flash E<sup>2</sup>PROM cell according to the present invention;

25

Figures 12A and 12B show a circuit diagram and cross-sectional diagram, respectively, of the flash E<sup>2</sup>PROM cell of the preferred embodiment configured in a cross point array architecture;

30

Figure 13 shows a simulation of programming time as a function of channel length for different p<sup>+</sup> doping concentrations for the preferred embodiment of Figure 9;

35

Figure 14 shows simulated and measured programming characteristics for the preferred embodiment of Figure 9; and

Figure 15 shows the experimental effect of the  $p^+/n^+$  junction on the disturb characteristics of a half-selected programmed and erased E<sup>2</sup>PROM cell according to the preferred embodiment of Figure 9.

5

#### Detailed Description of the Prior Art ETOX<sup>TM</sup> Memory Cell

Turning to Figure 1, a flash E<sup>2</sup>PROM cell is shown (ETOX<sup>TM</sup> cell by Intel Corp.), having a floating gate 1 and a control gate 3 stacked vertically above the MOPSET channel 5 in substrate 6 intermediate the source 7 and drain 9. A tunnel oxide 11 is provided intermediate the floating gate 1 and channel 5, and an inter-poly oxide 13 is provided intermediate the floating gate 1 and control gate 3.

In order to program the ETOX<sup>TM</sup> flash memory cell of Figure 1, a voltage of +12V is applied to the control gate 3, the source 7 is grounded, and approximately +6V is applied between source 7 and drain 9 as shown in Figure 2. As discussed above, the control gate 3 is capacitively coupled to the floating gate 1.

The source-drain voltage (+6V) generates hot electrons that are swept across the channel 5 from source 7 to drain 9. These hot electrons collide with atoms in the channel 5 thereby creating additional hot electrons. The high voltage (+12V) on control gate 3 attracts the hot electrons across the thin tunnel oxide 11 where they accumulate on the floating gate 1. When enough electrons have accumulated, the cell switches from its "1" (erased) state to its "0" (programmed) state.

As the floating gate 1 becomes fully charged, the current across the oxide 11 reduces almost to zero as a result of the electric field in the oxide 11 becoming

repulsive to additional electron injection from the high electric field region at the drain 9. Therefore, electron injection during programming is a self limiting process.

5

The voltage applied to the drain 9 does not, in general, have a significant effect on the programmed threshold voltage of the EEPROM cell. However, it does affect the speed of programming of the device.

10

To erase a flash memory cell, Fowler-Nordheim tunnelling is used to remove electrons from the floating gate 1. More particularly, by floating the drain 9, grounding the control gate 3 and applying +12V to the source 7, an electric field is generated across the thin tunnel oxide 11 between the floating gate 1 and the source 7. This electric field attracts electrons off of the floating gate 1 toward the source 7, as shown in Figure 3.

20

When all of the excess electrons have been removed from the floating gate 1, the threshold voltage ( $V_{th}$ ) returns to its initial value thereby resetting the cell to a "1" state.

25

During reading, the prior art ETOX™ memory cell operates like an ordinary transistor. When the flash memory is read, address inputs select specific transistors within the memory cell array. Supply-voltage levels are applied to the control gate 3 and drain 9, while the source 7 is grounded as shown in Figure 4. By sensing the amount of current that is conducted in the channel 5 between the source 7 and the drain 9, it is possible to determine if the cell has been programmed or erased, as discussed above.

35

Detailed Description of a

First Embodiment of Flash E<sup>2</sup>PROM Cell

Turning now to Figure 5, a flash E<sup>2</sup>PROM cell is shown for achieving simultaneous reduction of programming time and operating voltages over the prior art cell of Figure 1.

The cell according to the present invention comprises a substrate 15, drain and source regions 17 and 19, respectively, a floating gate 21 and control gate 23, in the usual manner. Tunnel and inter-poly dielectric layers 22 and 24 are also provided in the usual manner. However, in contrast with the prior art ETOX<sup>TM</sup> flash memory cell, additional highly doped regions 25 and 27 are provided adjoining the drain and/or source regions 17 and/or 19, respectively. The additional highly doped regions 25 and 27 (occasionally referred to herein as pocket implants) extend along a portion of the width of the cell structure identified as the program section 29, the remaining sense section 31 being provided for the read operation of the cell, as discussed in greater detail below.

In the illustrated embodiment, the substrate is made of p-type material, the drain and source regions are fabricated from n<sup>+</sup>-type material, and the additional highly doped regions 25 and 27 comprise p<sup>+</sup>-type regions.

According to the present invention, programming is accomplished via low voltage Zener and/or avalanche breakdown at the reverse biased p<sup>+</sup>/n<sup>+</sup> junctions in the program section 29. High doping concentrations at these junctions results in a narrow depletion width so that only a small voltage is required to generate a sufficiently strong electric field to create hot electrons for storage on the floating gate 21.

The breakdown mechanism which dominates and the applied voltage required for breakdown depend on the doping of the  $p^+$  regions. Computer simulation results have shown that  $p^+$  doping concentrations in the range of  $1 \times 10^{18} \text{cm}^{-3}$  to  $5 \times 10^{19} \text{cm}^{-3}$  are suitable for the implementation of this structure.

As indicated above, the cell is programmed and read from the program and sense sections 29 and 31, respectively, and erased by Fowler-Nordheim tunnelling of electrons from the floating gate 21 to the source 19.

With reference to Figure 6, a schematic representation of the programming mode of operation is provided. In order to program a logic "0" into the cell, sufficient electrons must be stored on the floating gate 21 to raise the transistor threshold voltage ( $V_{th}$ ) so that the normal read voltage (e.g. 5 Volts) applied to control gate 23 is insufficient to turn the transistor on.

Accordingly, low positive voltages (less than 5V) are applied to the drain 17 and source 19 while a high programming voltage (in the range of 10 to 15 Volts) is applied to control gate 23. The substrate 15 is grounded. Low voltage Zener and/or avalanche breakdown at the reversed biased  $p^+/n^+$  junctions, results in the generation of hot electrons which are drawn to the floating gate 21 as a result of the high voltage applied to control gate 23.

In contrast with the prior art ETOX™ memory cell, the generation of hot electrons necessary to program the flash memory cell of the present invention does not require a large drain-to-source current flow. This is a result of the high electric field (e.g.  $10^6 \text{V/cm}$ ) generated in the narrow depletion regions 26 associated with the heavily doped  $p^+/n^+$  junctions.

Figure 7 schematically shows the read mode operation of the flash memory cell according to the present invention. In operation, a low read voltage (e.g. +5 Volts) is applied to control gate 23; a voltage lower than the breakdown voltage of the  $p^+/n^+$  junction is applied to the drain 17 (e.g. 13 volts), and the source 19 and substrate 15 are grounded. Operation in the read mode is essentially the same as discussed above with respect to the ETOX™ prior art memory cell. In particular, in an erased cell, the voltage applied to control gate 23 is sufficient to overcome the transistor turn-on threshold voltage ( $V_{th}$ ), and the drain-to-source current is detected in the sense section 31 by sense amplifier circuitry (not shown) and translated into a logic "1". Conversely, in a programmed cell, the added electrons stored on the floating gate 21 raise the transistor turn-on threshold voltage ( $V_{th}$ ) so that the read voltage applied to control gate 23 is insufficient to turn on the transistor. The absence of current is detected in the sense section 31 as a logic "0".

Erase is performed by applying a high negative voltage (in the range of -12 to -15 Volts) to the control gate 23, applying a voltage, lower than the breakdown voltage of the  $p^+/n^+$  junction (e.g. +3.5 volts), to the source 17, while grounding the substrate 15 and leaving the drain in an open circuit condition, as shown in Figure 8. The voltage difference between the control gate 23 and the source 19, creates a strong electric field, across the thin oxide 22 between the floating gate 21 and the source 19, which in turn causes the removal of electrons from the floating gate 21 by Fowler-Nordheim tunnelling.

In summary, the flash  $E^2$ PROM memory cell according to the present invention, with appropriate  $p^+$  dopings and a short gate length, results in extremely fast programming

of the cell (in the order of nanoseconds) at low supply voltages, making the cell particularly applicable in the field of portable electronics.

5     Description of a Preferred Embodiment of Flash E<sup>2</sup>PROM Cell

10     The preferred cell structure according to the present invention uses a single p<sup>+</sup>/n<sup>+</sup> junction in the programming section of the drain region of the transistor, as illustrated in Figures 9 and 10. This embodiment offers the advantage that the erasing voltage is not limited by the p<sup>+</sup>/n<sup>+</sup> junction breakdown voltage, whereby the erasing speed can be improved by applying higher voltages to the source. The cell is programmed from the drain side of the program section, erased from the source side, and read from the sense section of the cell, according to the principles discussed above in connection with Figures 5-8.

20

25     An improved method of reading the preferred embodiment of flash E<sup>2</sup>PROM cell shown in Figures 9 and 10, comprises the steps of grounding the substrate 15 and drain 17, applying a positive logic level voltage to the control gate 23, and applying a voltage of no greater than logic level to the source region 19. The voltage applied to the control gate 23 allows current to flow through the channel region in the sense section, from the source region to the drain region in the event that little or no charge is stored on the floating gate.

30

35     This eliminates possible band to band tunnelling leakage current at the junction of the p<sup>+</sup> pocket implant 25 and the n<sup>+</sup> drain 17, thus reducing the probability of sensing error on the stored data. The voltage applied to the source region 19 can be higher than the breakdown voltage (typically 3V or less) of the p<sup>+</sup> pocket implant/

n<sup>+</sup> drain junction, resulting in a higher current flow through the channel during read operation. This will allow faster read time and lower sensing error rate.

## 5     Process Flow for Fabricating Flash E<sup>2</sup>PROM

10     A detailed description of the flash EEPROM cell fabrication process is given below. All of the steps described are optimized to achieve the appropriate specifications for the flash cells according to the present invention. The wafers used in the fabrication process are p-type (boron doped) with resistivity of 1-1.5 Ohm-cm and orientation <100>. The steps are as follows, with reference to Figures 11(a) through 11(r):

15

1.            Grow field oxide on the silicon wafers, to a target oxide thickness of ~5000 - 7000Å.
2.            Define the device active areas using photolithography (mask #1). The cross-section after this step is shown in Figure 11(a).
3.            Etch the oxide to form the device active area.
- 25    4.            Remove the photoresist. The cross-section after this step is shown in Figure 11(b).
5.            Grow the tunnelling or gate oxide 22, to a target oxide thickness of 100Å.
- 30    6.            Implant boron (energy E=30 keV, dose QI=4x10<sup>12</sup> cm<sup>-2</sup>) for threshold voltage adjustment. The cross-section after this step is shown in Figure 11(c).
- 35    7.            Deposit amorphous silicon by low pressure chemical vapor deposition (LPCVD), the target



thickness is 3600 Å. The reason for using amorphous silicon is that it results in a smooth surface which is important for the interpoly dielectric growth discussed in greater detail below.

5

8. Oxidize the amorphous silicon layer, to a target oxide thickness of ~ 200 Å. During this step, the amorphous silicon is converted to polysilicon. The thin oxide is used as a screen oxide during the subsequent ion implantation of phosphorus.
9. Implant phosphorus (energy E=30 keV, dose  $QI=8 \times 10^{15} \text{ cm}^{-2}$ ). This ion implantation step is used for setting the doping of the polysilicon. The cross-section after this step is shown in Figure 11(d).
10. Perform a rapid thermal anneal to activate the phosphorus and redistribute in the polysilicon after ion implantation.
11. Define the first polysilicon layer using photolithography (mask #2). The cross-section after this step is shown in Figure 11(e).
12. Etch the thin screen oxide and polysilicon, remove the photoresist and etch the remaining areas of thin oxide.
13. The oxide-nitride-oxide (ONO) interpoly dielectric is grown in this step. The target equivalent thickness is 200 Å. The cross-section after this step is shown in Figure 11(f).

10

15

20

25

30

35

14. Deposit amorphous silicon by LPCVD to form the control gate 23. The target thickness is 3600 Å.
- 5 15. Oxidize the deposited amorphous silicon layer, to a target oxide thickness of ~200 Å. During this step, the amorphous silicon is converted to polysilicon. The thin oxide is used as a screen oxide during the subsequent ion  
10 implantation of phosphorus.
16. Implant phosphorus (energy E=30 keV, dose  $QI=8 \times 10^{15} \text{ cm}^{-2}$ ). This ion implantation step is used for setting the doping of the polysilicon.  
15 The cross-section after this step is shown in Figure 11(g).
17. Deposit LPCVD oxide. The target thickness is ~3500 Å. This oxide serves as a mask in  
20 Reactive Ion Etching (RIE) of the polysilicon layers.
18. Perform a rapid thermal anneal step to activate the phosphorus implanted in the polysilicon and  
25 densify the LPCVD oxide.
19. Define the stacked gate area over LPCVD oxide using photolithography (mask #3). The cross-section after this step is shown in Figure  
30 11(h).
20. Etch the LPCVD oxide using RIE.
21. Remove the photoresist.
- 35 22. Perform a RIE of the second polysilicon layer. The control gate 23 is formed in this step.

23. Perform a RIE of the undesired areas of  
interpoly oxide 24 and a RIE of the undesired  
areas of the first polysilicon layer. The  
floating gate is formed in this step. The  
remaining undesired areas of the first  
polysilicon are removed from the source and  
drain regions. The cross-section after this  
step is shown in Figure 11(i).
24. Perform a dry oxidation to ensure that the  
floating gate is encapsulated by a high quality  
dielectric layer.
25. Deposit thin LPCVD oxide. The target oxide  
thickness is  $\sim 500 \text{ \AA}$ . This oxide is used as a  
screen layer during the ion implantation to  
form the substrate contacts and the  $P^+$  region of  
the Zener injector.
26. Define the windows for boron implantation using  
photolithography (mask #4). This  
photolithography process step opens a window  
through which the  $p^+$  region of the Zener  
junction and the substrate contacts are formed.  
This step defines the program section of the  
cell.
27. Implant boron (energy  $E=40\text{keV}$ , dose  
 $QI=2 \times 10^{15} \text{cm}^{-2}$ ). This ion implantation step is  
used for implanting the  $p^+$  regions of the Zener  
junction in order to form a junction at the  
drain side of the gate and also forms the  
substrate contacts. The cross-section after  
this step is shown in Figure 11(j).
28. Remove the photoresist.

29. Etch the thin screen oxide. A wet oxide etch is used in this step.
- 5 30. Deposit LPCVD oxide. The target oxide thickness is ~ 3500 Å. This oxide layer is used for the formation of the side wall spacers (SWS).
- 10 31. Anneal the wafers to densify the LPCVD oxide and to drive-in boron after the implantation. The cross-section after this step is shown in Figure 11(k).
- 15 32. Etch the deposited LPCVD oxide layer. The side wall spacers (SWS) are formed in this step. Perform a RIE of oxide in the process. The SWS are very important for the location of the p<sup>+</sup>/n<sup>+</sup> junction underneath the gate oxide 22. Experiments show that for this particular  
20 stacked gate structure and this range of LPCVD oxide thickness (~ 3500 Å), the SWS width should be approximately 58% of the LPCVD oxide layer thickness. The cross-section after this step is shown in Figure 11(l).
- 25 33. Deposit thin LPCVD oxide. The target oxide thickness is ~ 500 Å. This oxide is used as a screen layer during the subsequent ion  
30 implantation which forms the source and the drain regions.
- 35 34. Define windows for phosphorus implantation using photolithography (mask #5). This photolithography process step defines windows through which n<sup>+</sup> source and drain regions are implanted. In this step the photoresist covers p<sup>+</sup> diffusion regions designed to serve as

contacts to the substrate. The windows are open only over the device active area.

- 5           35.       Implant phosphorus (energy  $E=80$  keV, dose  $QI=8 \times 10^{15}$  cm<sup>-2</sup>). This ion implantation step is used for implanting the n<sup>+</sup> source and drain regions 19 and 17. The cross-section after this step is shown in Figure 11(m).
- 10          36.       Remove the photoresist. and etch the thin screen oxide.
- 15          37.       Deposit LPCVD oxide. The target oxide thickness is ~ 8000 Å. This oxide is used as an isolation layer between the devices and the metal layer.
- 20          38.       Perform a thermal drive-in to form the metallurgical junction of the p<sup>+</sup>/n<sup>+</sup> injector to be directly underneath the floating gate electrode (on the drain side).
- 25          39.       Define the contact windows before aluminum deposition using photolithography (mask #6). This photolithography process step defines metal to diffusion and polysilicon contacts.
- 30          40.       Etch oxide to define contact windows. The cross-section after this step is shown in Figure 11(n).
41.       Remove the photoresist.
- 35          42.       Sputter aluminum on the wafers. The cross-section after this step is shown in Figure 11(o).
43.       Define the aluminum using photolithography

(mask #7). The cross-section after this step is shown in Figure 11(p).

44. Etch the aluminum.

45. Remove the photoresist.

46. Metal sintering is done in forming gas. The final cross-section of the cell's program section and sense section are shown in Figures 11(q) and in Figure 11(r) respectively.

It should be noted that the p<sup>+</sup> implant step used to form the injector in the fabrication of this device is similar to the step used in the implementation of the n-LDD implant in a standard CMOS/BiCMOS process flow.

#### Application of the Flash E<sup>2</sup>PROM of the Present Invention

The flash E<sup>2</sup>PROM cell of the preferred embodiment can be configured in a cross point array architecture, as shown in Figures 12A and 12B. The p<sup>+</sup> pocket implant does not necessarily result in a larger cell size since it can be accommodated within the minimum device width. However, the cell exhibits a lower read current. The p-buried layer 15A provides sufficient substrate current extraction to minimize interference with on-chip logic circuits.

#### Experimental and Test Results

The programming time of the flash E<sup>2</sup>PROM of the present invention is dependent on the p<sup>+</sup> region doping levels, and has been found by computer simulation to be at least an order of magnitude less than that of an

equivalent dimension conventional flash E<sup>2</sup>PROM cell operating under the same conditions, as shown in Figure 13. The decrease in programming time with decreased channel length is mainly due to the reduction in gate area since the channel length has no effect on the generation of hot electrons. A heavily doped p<sup>+</sup>/n<sup>+</sup> junction between the drain 17 and pocket implant 25 is preferred since breakdown occurs at lower voltages and higher electric field strength, resulting in faster programming speeds. With an appropriate p<sup>+</sup> doping concentration of  $>10^{19}\text{cm}^{-3}$  and a short channel length (e.g.  $\sim 0.5\text{ }\mu\text{m}$ ), the programming time is on the order of 150 ns for 3.3V operation.

Successful prototypes of the flash E<sup>2</sup>PROM memory cell in accordance with the preferred embodiment have been fabricated using a conventional CMOS process flow as discussed in detail above with reference to Figures 11(a) through 11(r). According to the successful prototypes, the device gate length was established as  $3\text{ }\mu\text{m}$ , the tunnel oxide 22 thickness was 100 Angstroms and the interpoly oxide 24 was 300 Angstroms. A heavy boron implant was used to form the p<sup>+</sup> region 25 with an effective doping concentration of approximately  $1.8 \times 10^{18}\text{cm}^{-3}$ .

The measured and simulated programming times plotted in Figure 14 are seen to be in good agreement. For a  $3\text{ }\mu\text{m}$  channel length with p<sup>+</sup> doping concentration of  $1.8 \times 10^{18}\text{cm}^{-3}$ , the programming time was measured to be  $6\text{ }\mu\text{s}$  at a drain bias of 6.5V which is an order of magnitude smaller than that of an equivalent dimension prior art conventional flash E<sup>2</sup>PROM cell operating under the same conditions. The drain-to-substrate current was limited to  $120\text{ }\mu\text{A}$  per  $\mu\text{m}$  of channel width, which is comparable to currents observed in conventional flash E<sup>2</sup>PROM cells. The erase time was measured at 100 milliseconds which is consistent with prior art device characteristics.

Shorter erase times may be obtained by increasing the coupling ratio between the control and floating gates. Furthermore, a measured read current of 100  $\mu\text{A}$  per  $\mu\text{m}$  of device width was obtained from the memory cell in the  
5 erased state with +3 Volts and +5 Volts applied to the drain 17 and control gate 23, respectively, which is also consistent with prior art operating characteristics.

The effect of the  $p^+/n^+$  junction on the disturbed characteristics of half-selected cells is illustrated in  
10 Figure 15. The drain disturbance for programmed cells becomes apparent if the bit-lines are raised high for more than one second. This represents a significant safety margin for sequential programming in a large  
15 array. The write/erase endurance characteristics of the cells were also measured to be larger than 10,000 cycles.

Other embodiments and modifications of the invention are possible. For example, as a further alternative to  
20 the preferred embodiment of Figure 9, the pocket  $p^+$  implant 25 may be located adjacent the source region 19, with minor modifications being necessary to the applied voltage in various modes of operation. Also, an improved version of these devices can be fabricated by using a  $p^+$   
25 buried layer in the substrate to collect the substrate current generated during the programming operation. All such modifications and alternatives are believed to be within the sphere and scope as defined by the claims appended hereto.



## WE CLAIM

1. In a flash E<sup>2</sup>PROM cell having n<sup>+</sup> source and drain regions disposed in a p-type substrate, a channel region  
5 intermediate to said source and drain regions, a tunnel dielectric layer overlying said channel region, a floating gate overlying said tunnel dielectric layer, an inter-poly dielectric layer overlying said floating gate and a control gate overlying said inter-poly dielectric  
10 layer, the improvement comprising a highly doped p<sup>+</sup> pocket implant adjoining one of said drain and source regions along a portion of the width of said cell for creating a metallurgical junction having narrow depletion width, said portion of the width of said cell defining a program  
15 section of said cell and the remaining width of said cell defining a sense section thereof.

2. The improvement of claim 1, wherein said p<sup>+</sup>-type material has a doping concentration in the range of from  
20  $1 \times 10^{18} \text{ cm}^{-3}$  to  $1 \times 10^{20} \text{ cm}^{-3}$ .

3. The improvement of claim 1 or 2, wherein said metallurgical junction is driven in so as to be  
25 positioned directly beneath said floating gate.

4. A method of programming the improved flash E<sup>2</sup>PROM cell of claim 1, comprising the steps of:  
a) grounding said substrate;  
b) applying a low positive voltage to said one of  
30 said drain and source regions adjoining said p<sup>+</sup> pocket implant, whereby said junction becomes reverse biased and one of either low voltage Zener or avalanche breakdown occurs resulting in generation of hot electrons; and  
c) applying a high positive voltage to said control  
35 gate, whereby said hot electrons are attracted to said floating gate.

5. A method of erasing the improved flash E<sup>2</sup>PROM cell of claim 1, comprising the steps of:

- a) grounding said substrate;
- b) applying a logic level voltage to the other of  
5 said drain and source regions;
- c) leaving said one of said drain and source regions adjoining the p<sup>+</sup>pocket implant in an open circuit condition; and
- d) applying a high negative voltage to said control  
10 gate, whereby a strong electric field is generated between said floating gate and said other of said drain and source regions resulting in removal of electrons from said floating gate to said other of said drain and source regions by Fowler-Nordheim tunnelling.

15 6. A method of reading the improved flash E<sup>2</sup>PROM cell of claim 1, comprising the steps of:

- a) grounding said substrate and said source region;
- b) applying a low voltage to said control gate; and
- 20 c) applying a voltage less than breakdown voltage of said junction to said drain region, whereby said voltage applied to the control gate causes current to flow through said channel region in said sense section from said drain region to said source region.

25 7. A method of reading the improved flash E<sup>2</sup>PROM cell of claim 1, comprising the steps of:

- a) grounding said substrate and said one of said  
30 drain and source regions adjoining said p<sup>+</sup>pocket implant;
- b) applying a positive logic level voltage to said control gate;
- c) applying a voltage of no greater than logic  
35 level to said other one of said source and drain regions, whereby said voltage applied to the control gate allows current to flow through said channel region in said sense section, from

said source region to said drain region in the event that little or no charge is stored on said floating gate.

5

8. In a flash E<sup>2</sup>PROM cell having n<sup>+</sup> source and drain regions disposed in a p-type substrate, a channel region intermediate to said source and drain regions, a tunnel dielectric layer overlying said channel region, a floating gate overlying said tunnel dielectric layer, an inter-poly dielectric layer overlying said floating gate and a control gate overlying said inter-poly dielectric layer, the improvement comprising a pair of highly doped p<sup>+</sup> pocket implants adjoining respective ones of said drain and source regions along a portion of the width of said cell for creating a pair of metallurgical junctions each having narrow depletion width, said portion of the width of said cell defining a program section of said cell and the remaining width of said cell defining a sense section thereof.

20

9. The improvement of claim 8, wherein said p<sup>+</sup>-type material has a doping concentration in the range of from  $1 \times 10^{18} \text{ cm}^{-3}$  to  $1 \times 10^{20} \text{ cm}^{-3}$ .

25

10. The improvement of claim 8 or 9, wherein said metallurgical junctions are driven in so as to be positioned directly beneath said floating gate.

30

11. A method of programming the improved flash E<sup>2</sup>PROM cell of claim 8, comprising the steps of:

a) grounding said substrate;

b) applying a low positive voltage to each of said drain and source regions, whereby said junctions become reverse biased and one of either low voltage Zener or avalanche breakdown occurs resulting in generation of hot electrons; and

35

c) applying a high positive voltage to said control gate, whereby said hot electrons are attracted to said floating gate.

- 5      12. A method of erasing the improved flash E<sup>2</sup>PROM cell of claim 8, comprising the steps of:
- a) grounding said substrate;
- b) applying a voltage less than breakdown voltage of said junction to one of said drain and source regions;
- 10        c) leaving the other of said drain and source regions in an open circuit condition; and
- d) applying a high negative voltage to said control gate, whereby a strong electric field is generated between said floating gate and said one of said drain and
- 15        source regions resulting in removal of electrons from said floating gate to said one of said drain and source regions by Fowler-Nordheim tunnelling.

- 20      13. A method of reading the improved flash E<sup>2</sup>PROM cell of claim 8, comprising the steps of:
- a) grounding said substrate and said source region;
- b) applying a low voltage to said control gate; and
- c) applying a voltage less than breakdown voltage of
- 25        said junction to said drain region, whereby said voltage applied to the control gate causes current to flow through said channel region in said sense section from said drain region to said source region.

- 30      14. A method of fabricating a flash E<sup>2</sup>PROM cell, comprising the steps of:
- a) growing a gate oxide layer on a p-type substrate;
- b) depositing amorphous silicon on said gate oxide layer;
- 35        c) oxidizing said amorphous silicon to create polysilicon;
- d) doping said polysilicon to create a floating gate

layer;

e) growing an interpoly dielectric layer over said floating gate layer;

f) depositing a second amorphous silicon layer on said interpoly dielectric layer;

g) oxidizing said second amorphous silicon layer to create a second polysilicon layer;

h) doping said second polysilicon layer to create a control gate layer;

i) masking and etching said gate oxide layer, floating gate layer, interpoly dielectric layer and control gate layer to form a defined stack of gate oxide, floating gate, interpoly dielectric and control gate, respectively;

j) doping said substrate adjacent one side of said stack with a p<sup>+</sup> impurity, thereby creating a pocket implant region along a program region forming a first portion of the width of said cell, a sense region forming the remainder of the width of said cell;

k) extending an oxide side wall spacer from both sides of said stack;

l) doping said substrate with an n<sup>+</sup> impurity, for creating source and drain regions, such that a narrow depletion width metallurgical p<sup>+</sup>/n<sup>+</sup> junction is created between said pocket implant region and said drain region within said program region of said cell upon application of a reverse bias voltage;

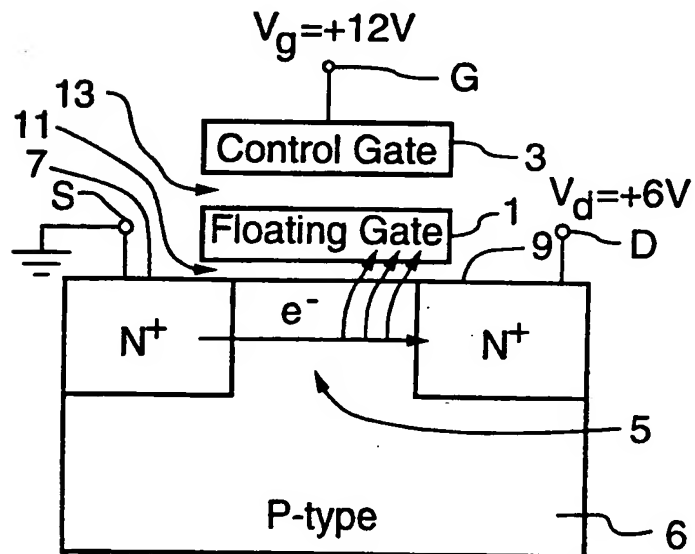
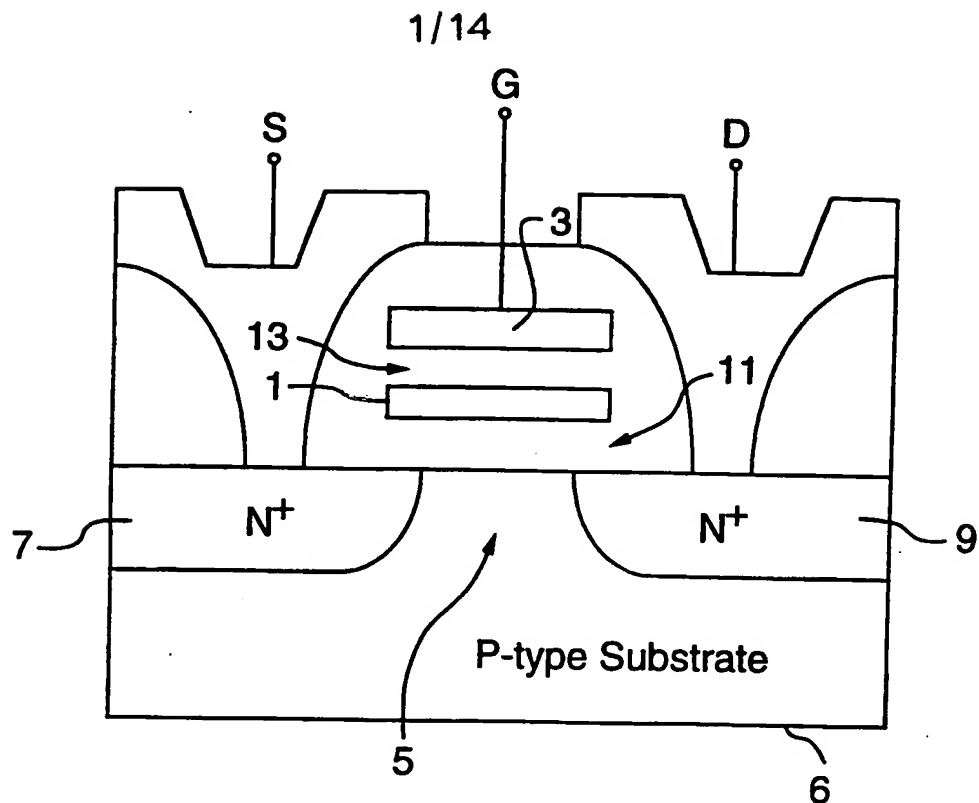
m) performing a thermal drive-in for locating said metallurgical p<sup>+</sup>/n<sup>+</sup> junction entirely beneath said stack;

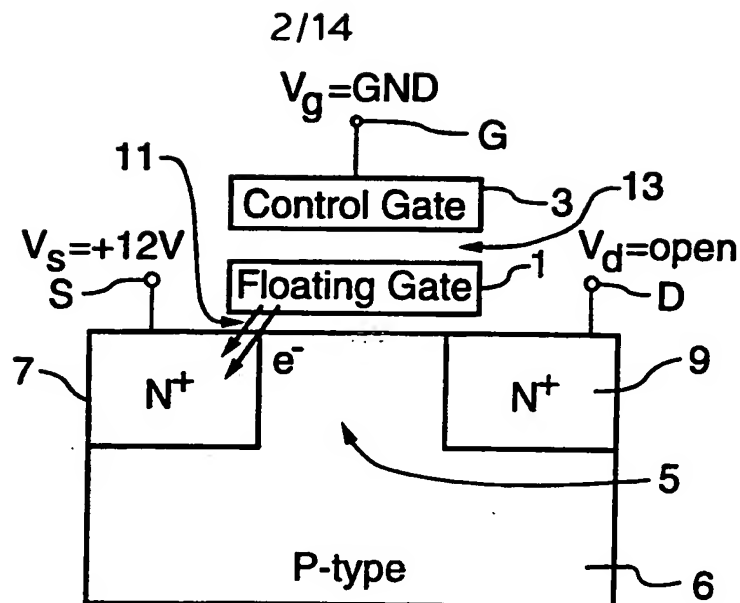
and

n) depositing and patterning metal contacts for said substrate, said control gate and said source and drain regions.

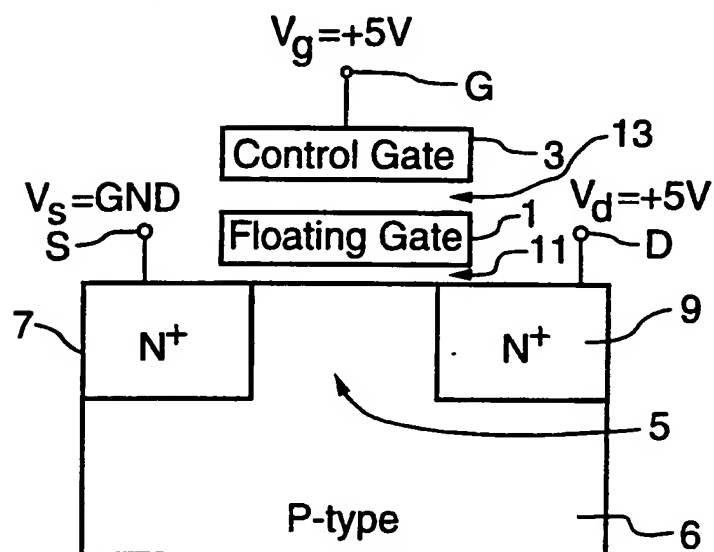
15. The method of claim 14, further comprising the step of doping a buried layer in said substrate with an impurity of said first type, for collecting substrate

current during programming of said cell.





**FIG.3 (PRIOR ART)**



**FIG.4 (PRIOR ART)**



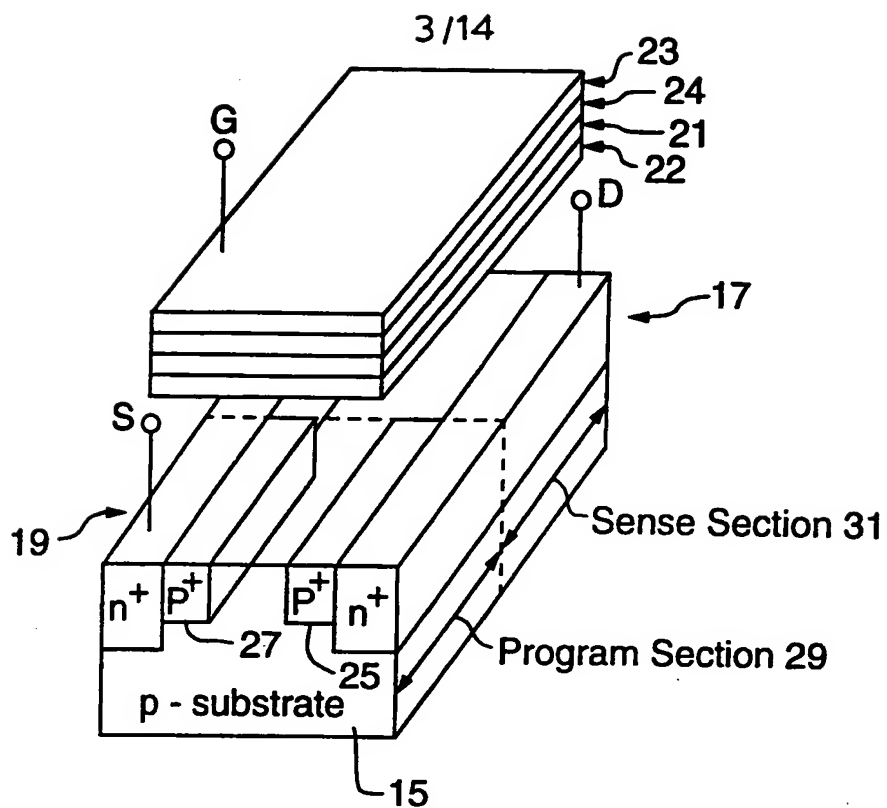


FIG. 5

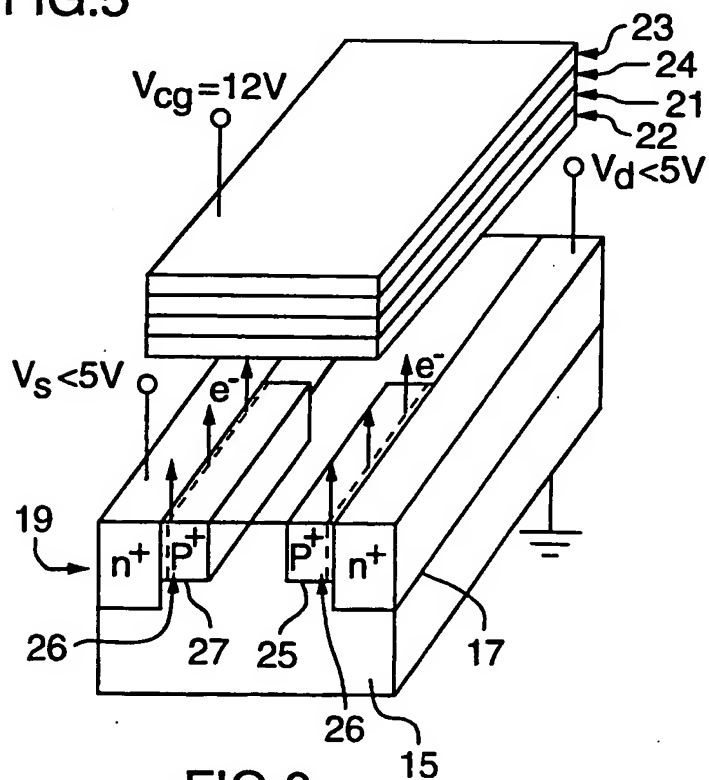


FIG. 6

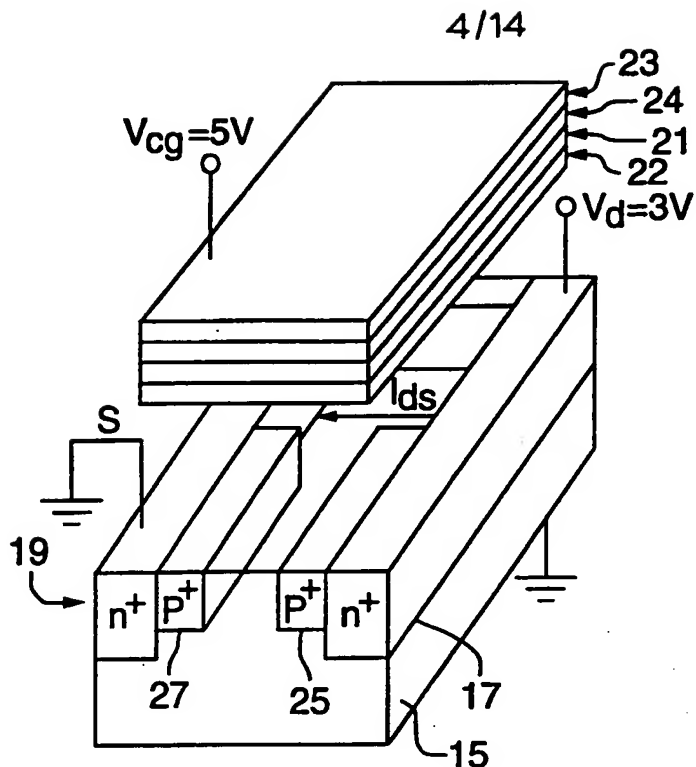


FIG.7

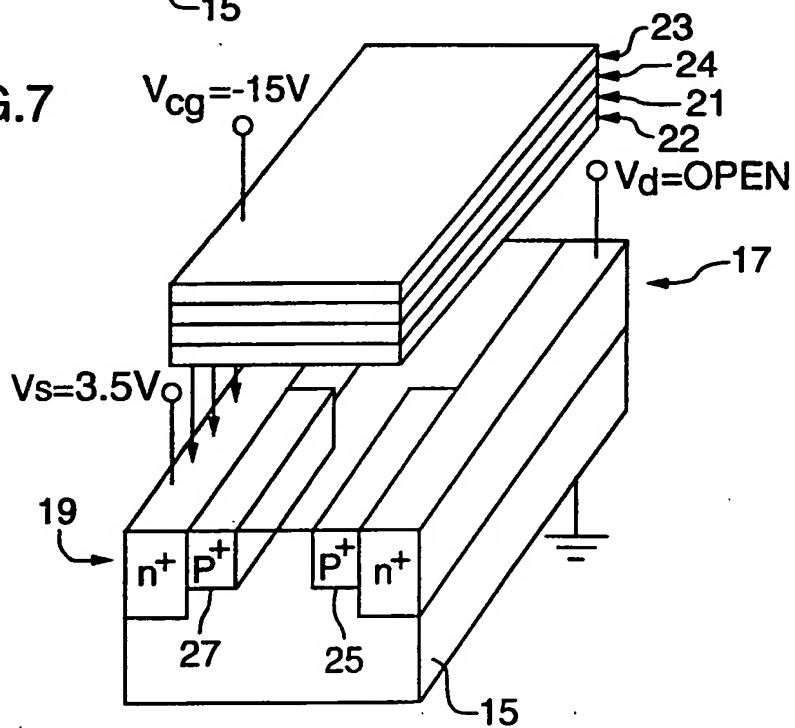


FIG.8

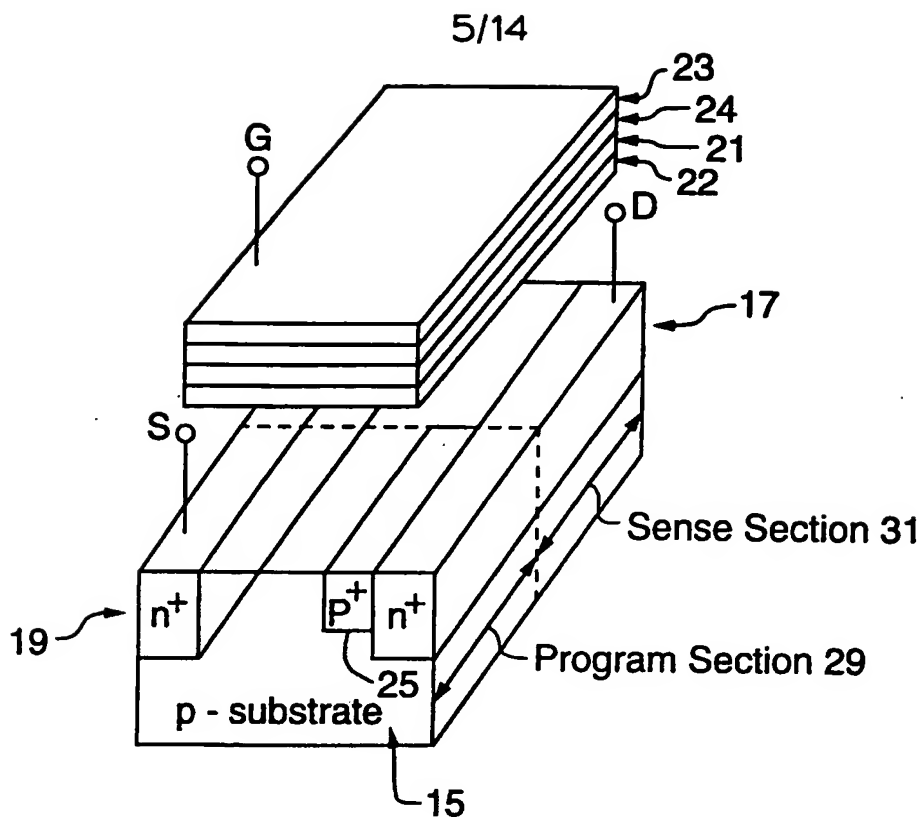


FIG.9

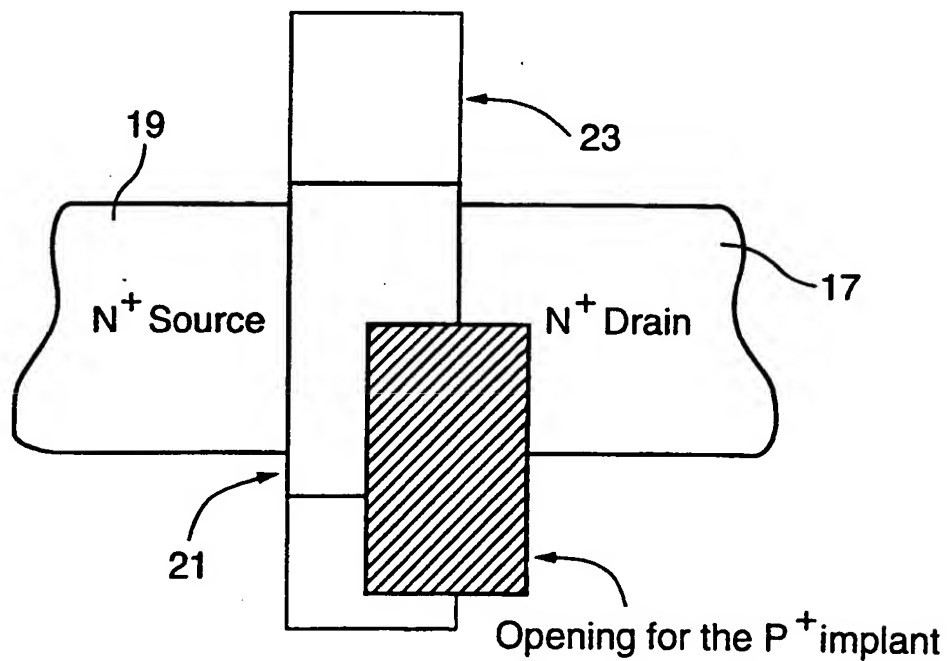


FIG.10

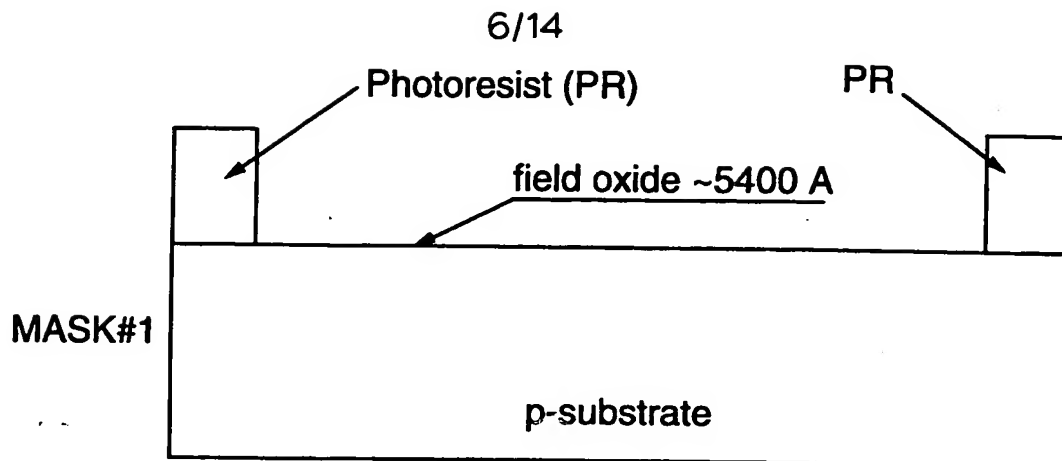


FIG.11A

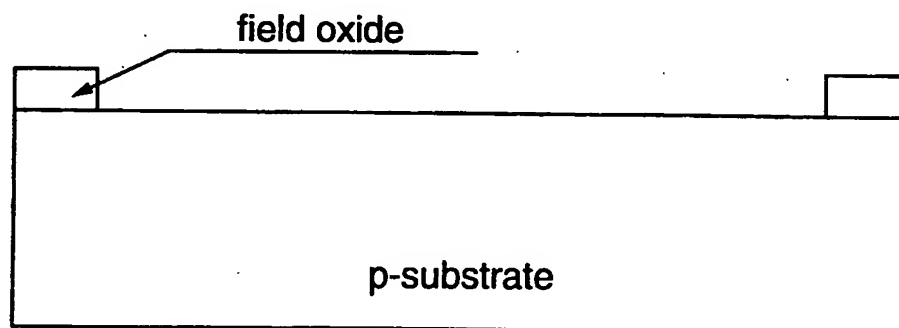


FIG.11B

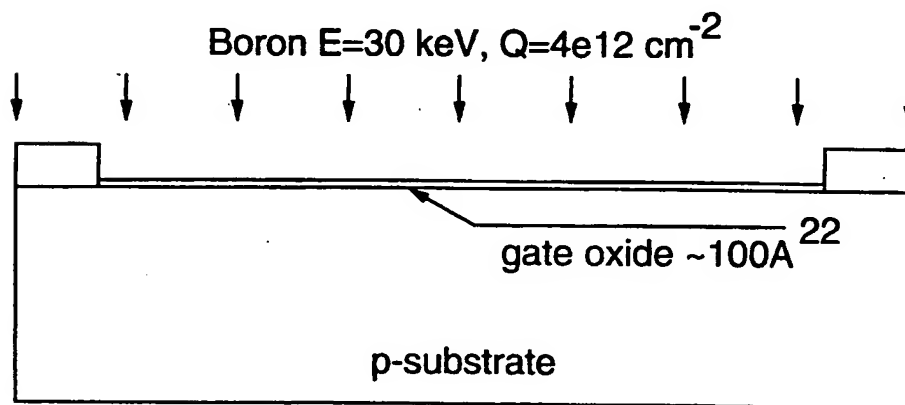


FIG.11C

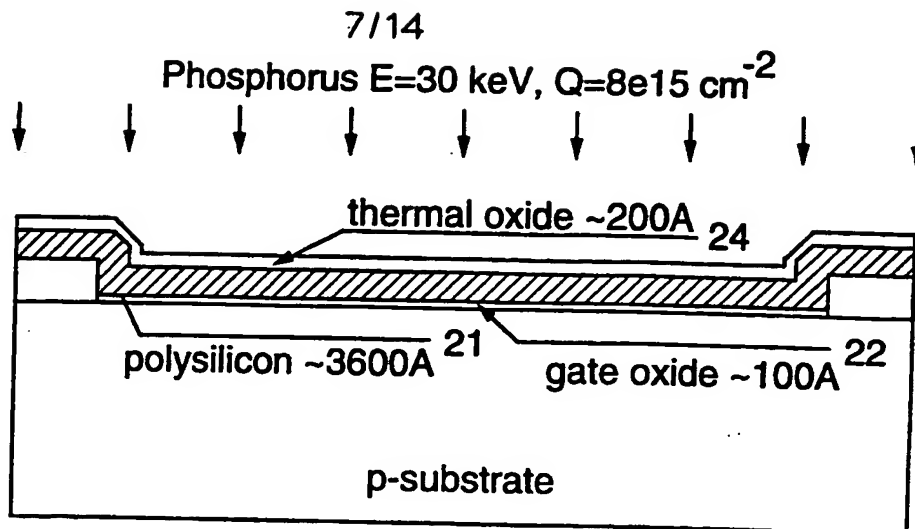


FIG.11D

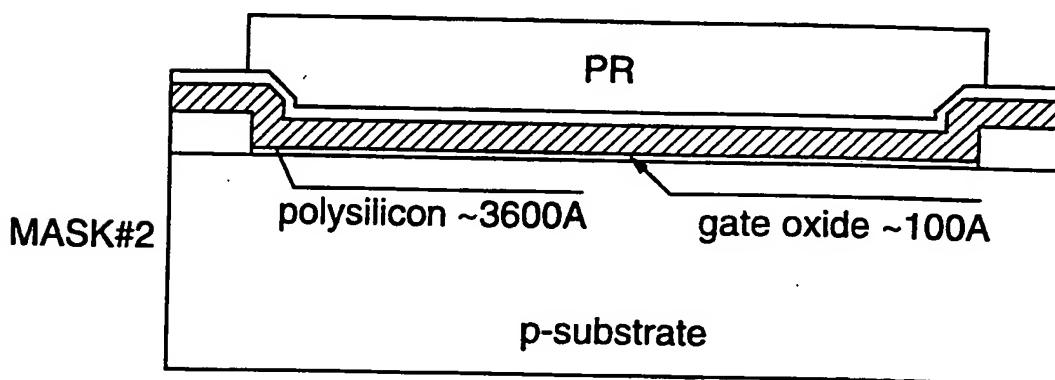


FIG.11E

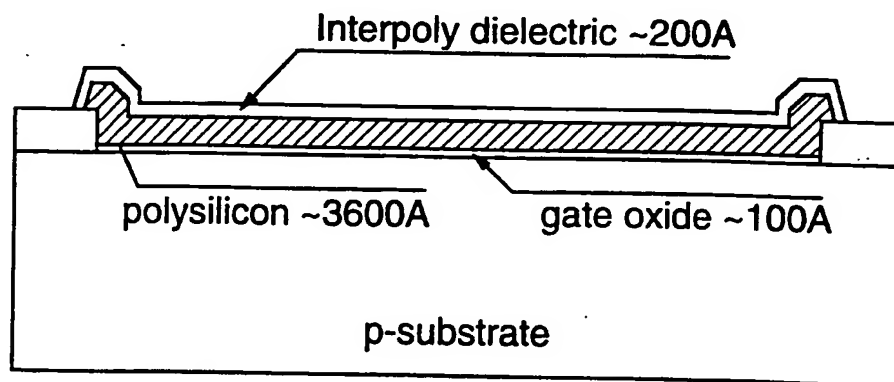


FIG.11F

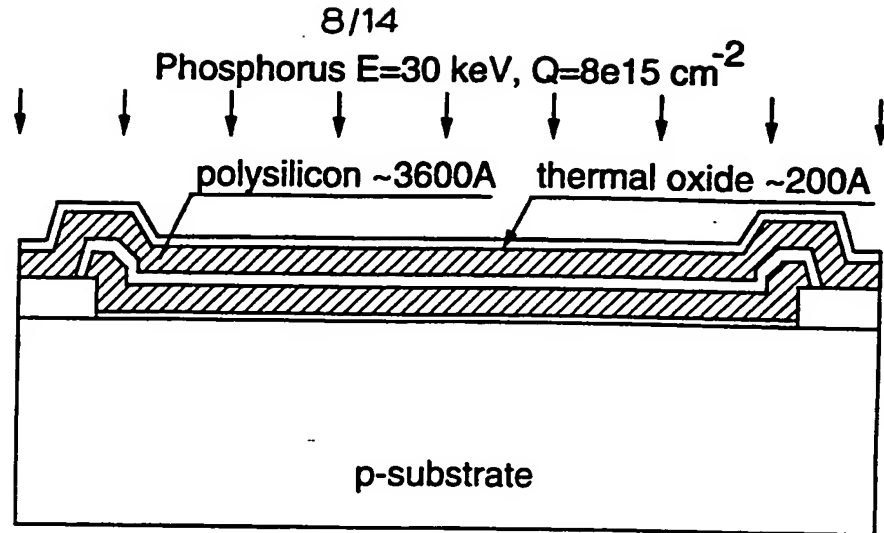


FIG.11G

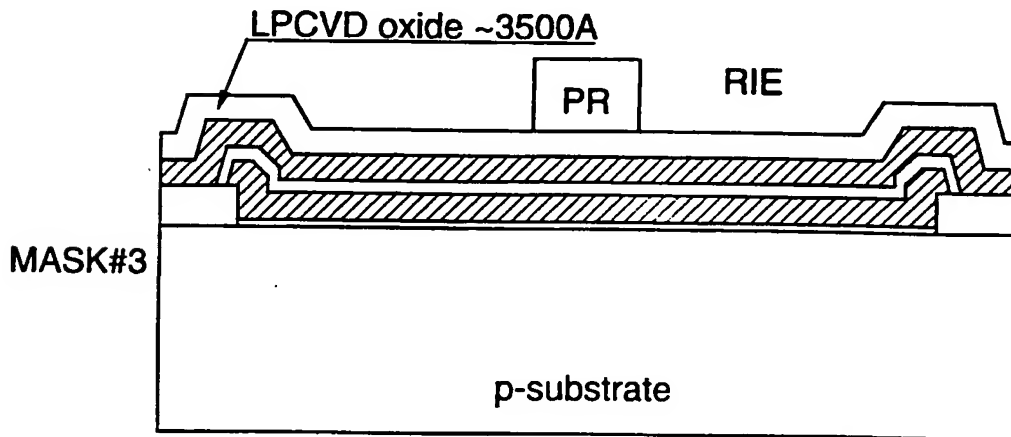


FIG.11H

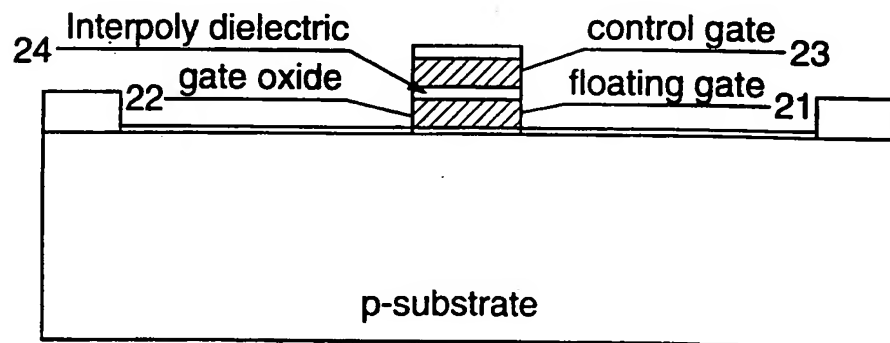


FIG.11I

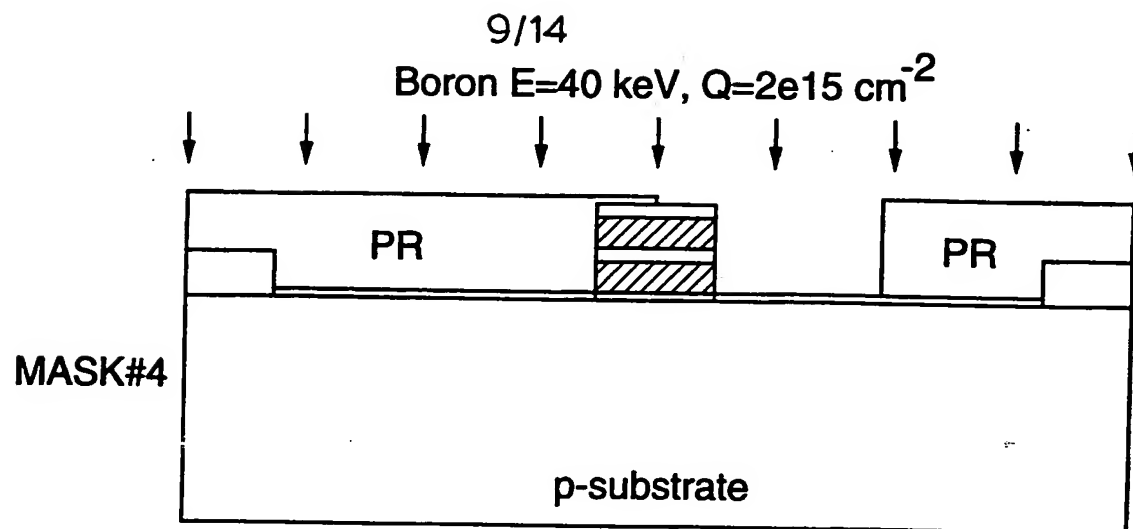


FIG.11J

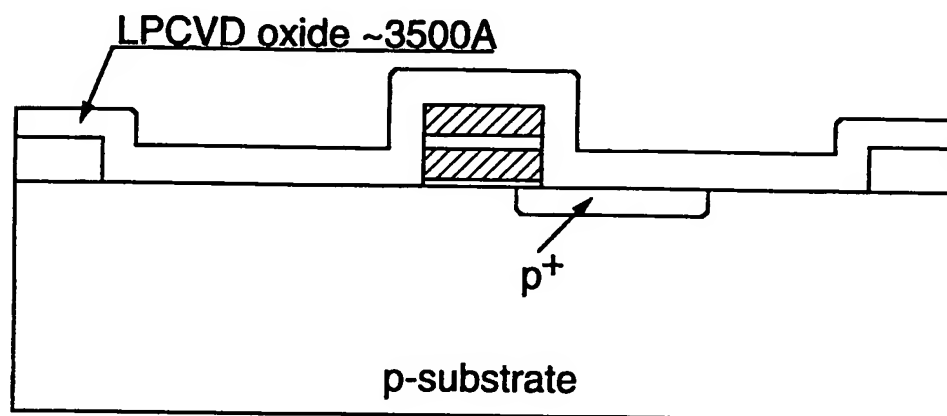


FIG.11K

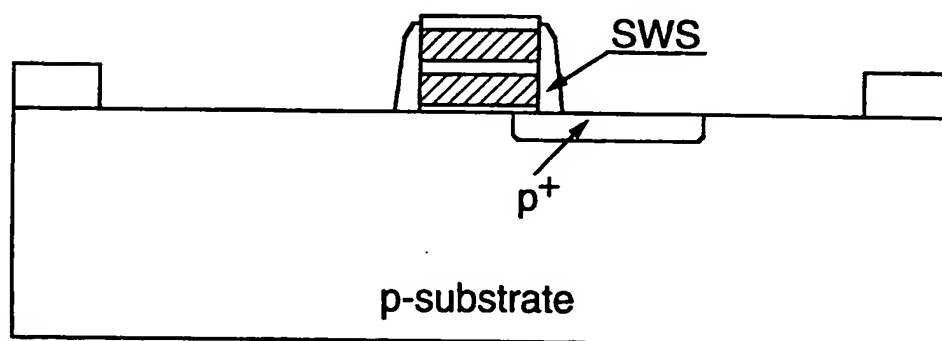


FIG.11L

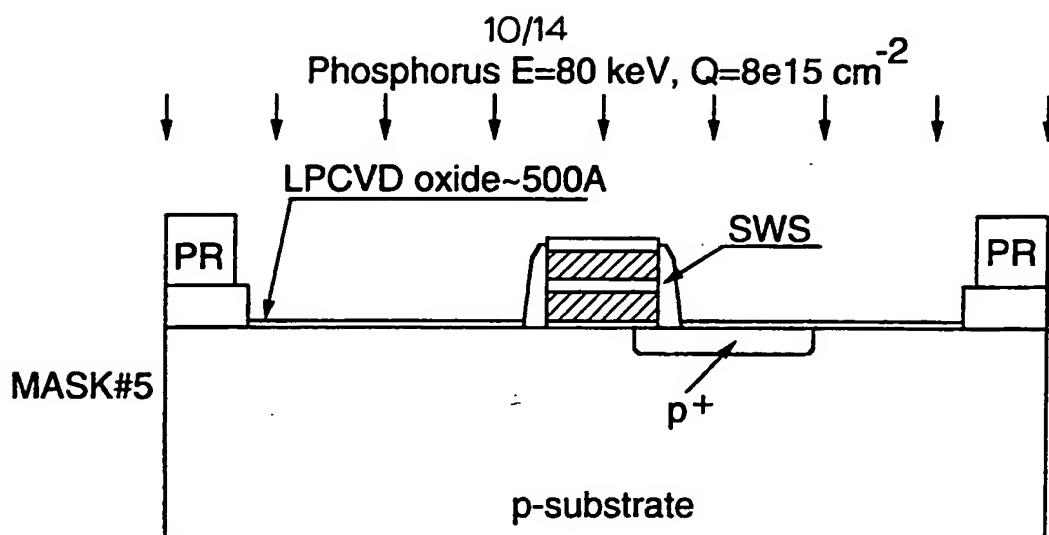


FIG.11M

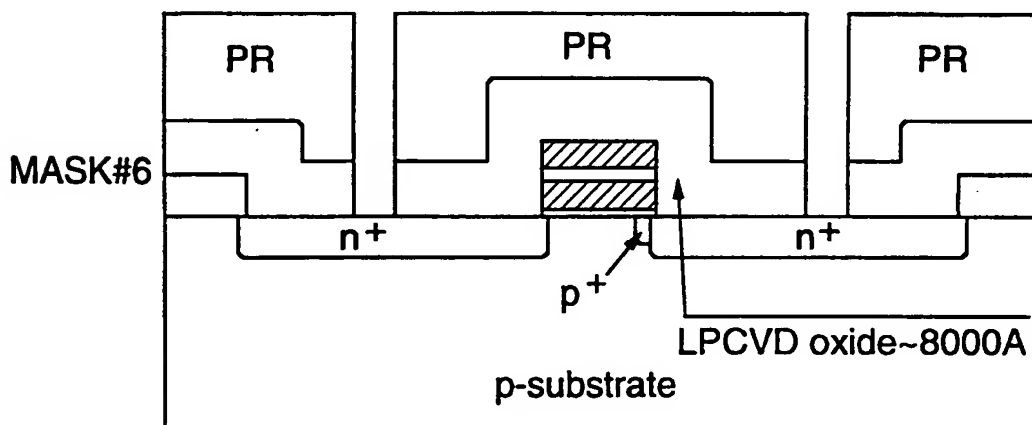


FIG.11N

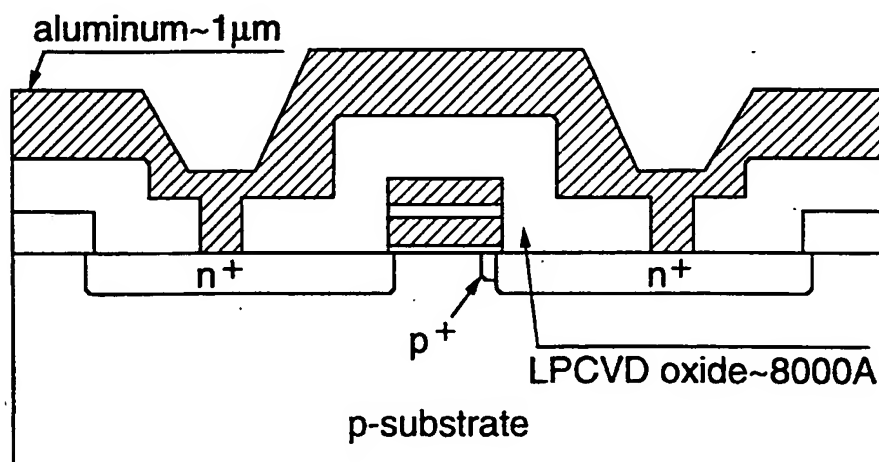


FIG.11.O



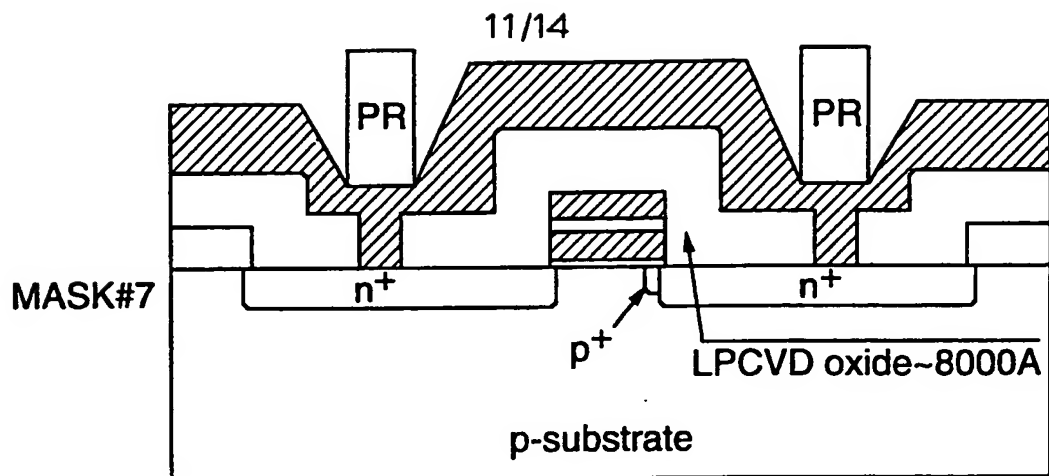


FIG.11P

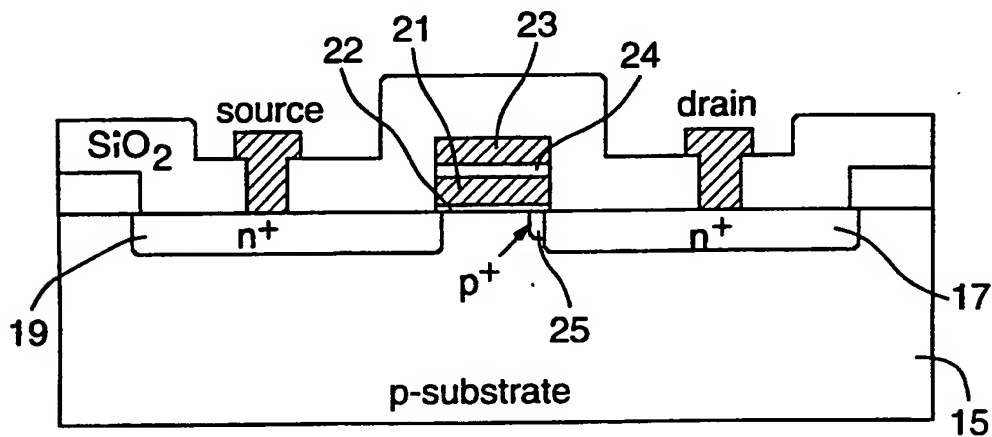


FIG.11Q

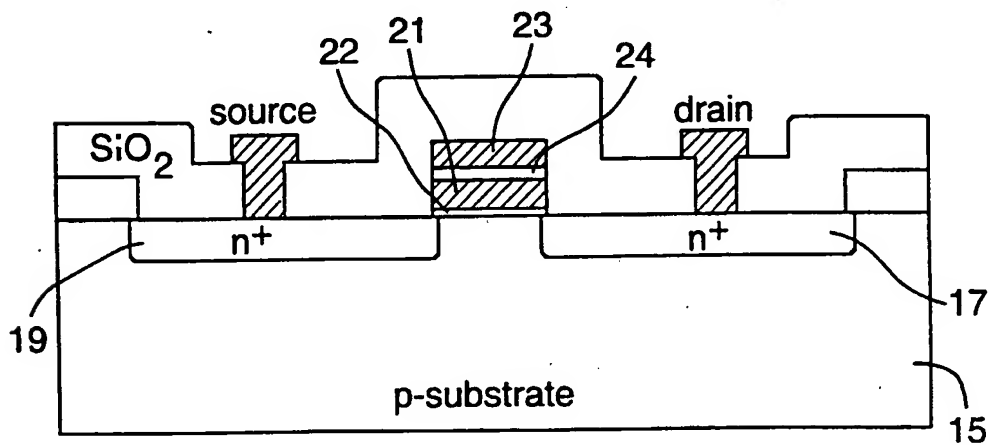


FIG.11R

12/14

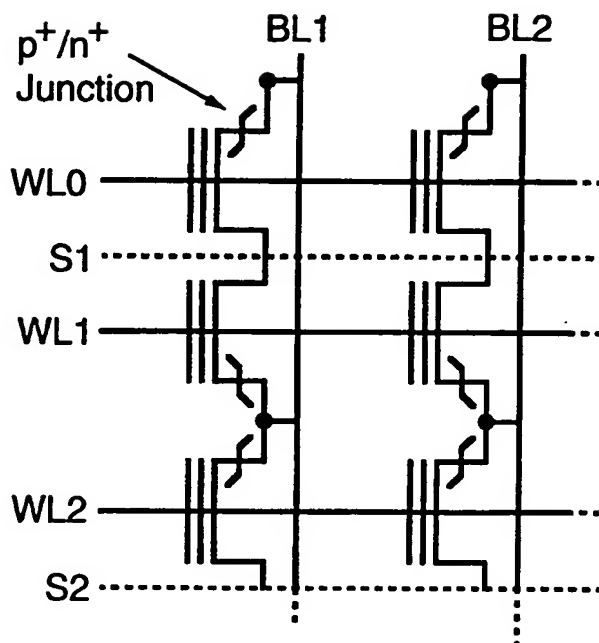


FIG. 12A

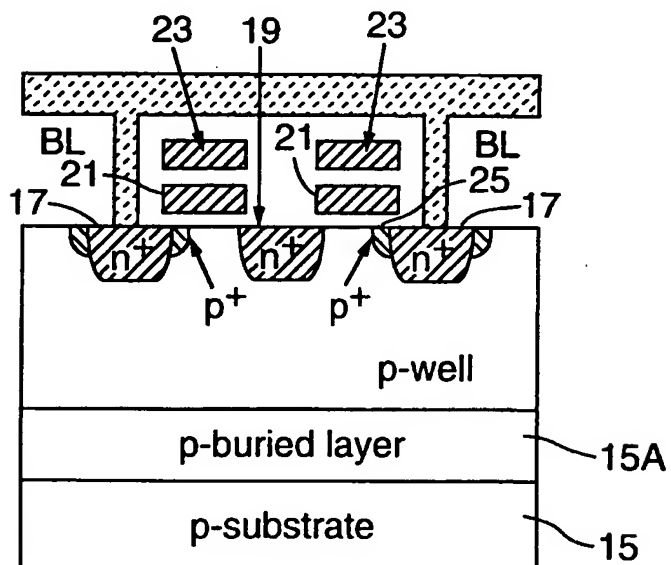


FIG. 12B

13/14

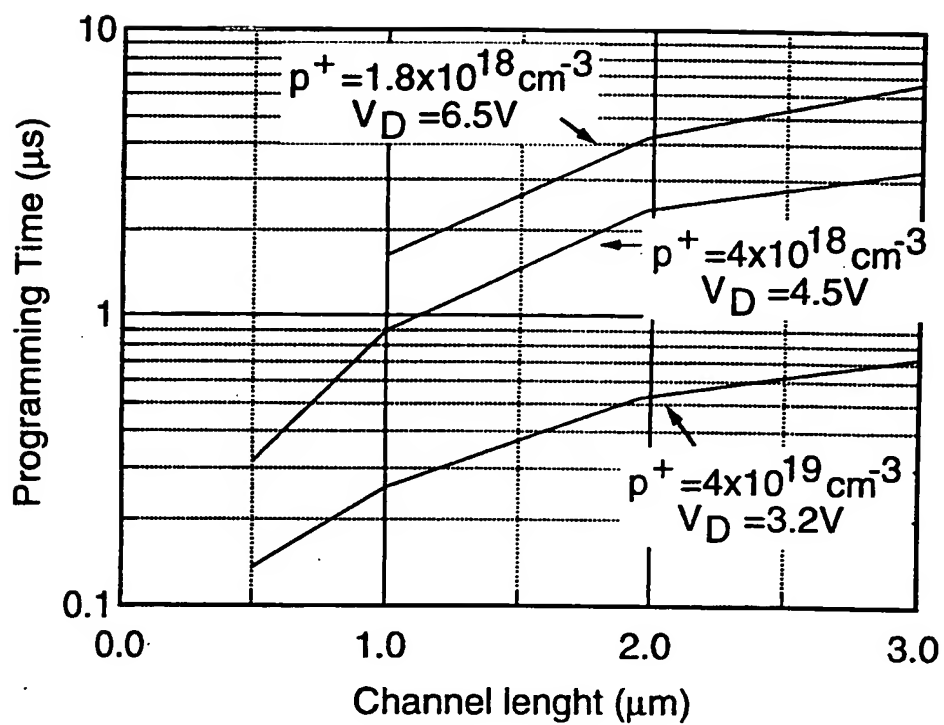


FIG.13

14/14

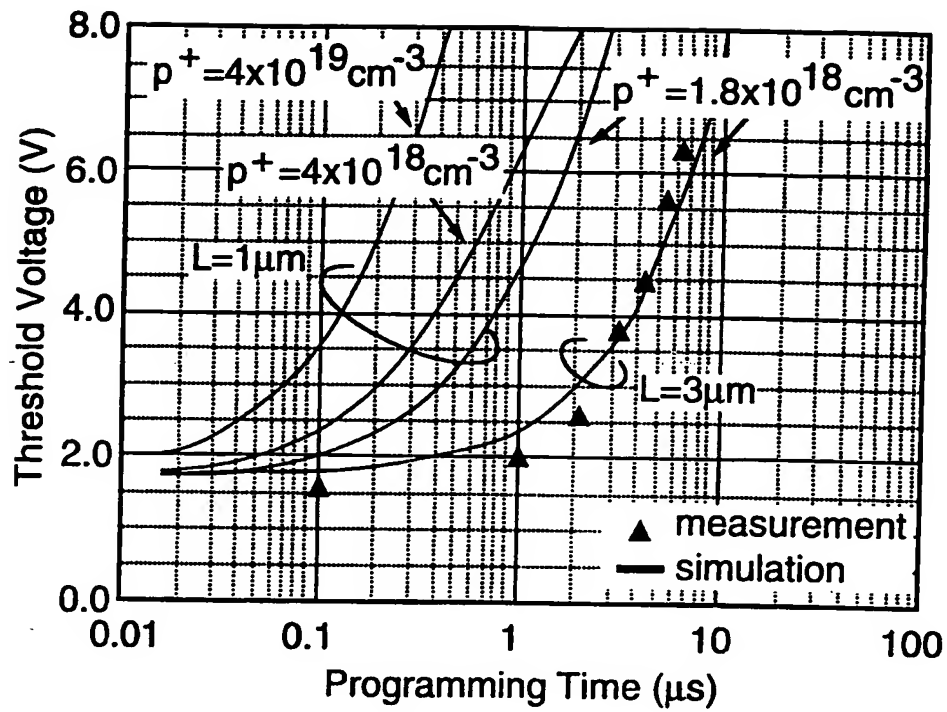


FIG. 14

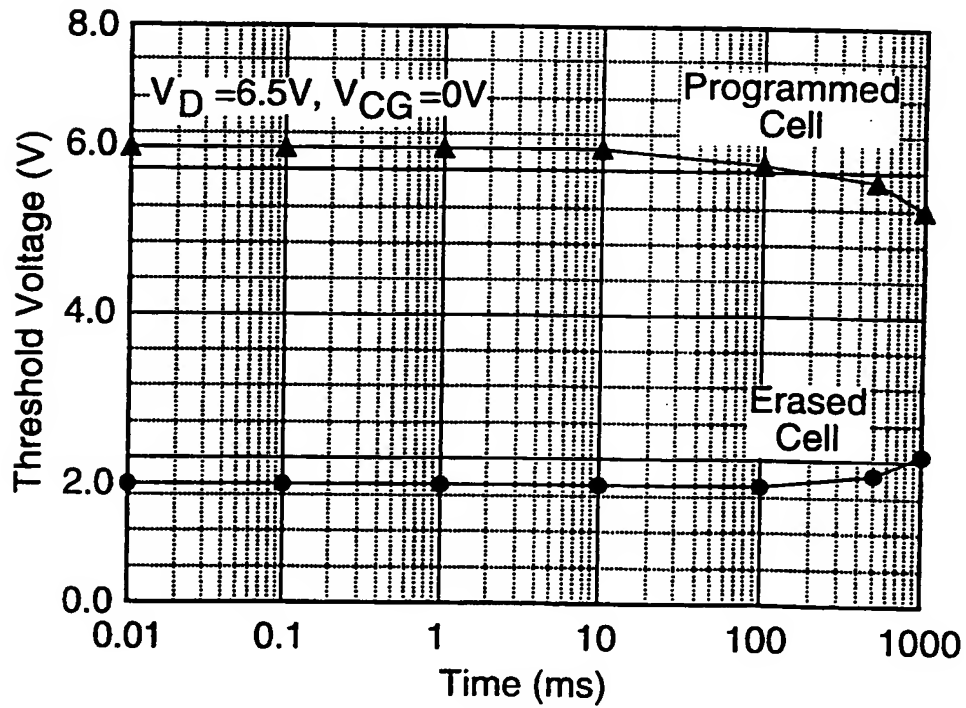


FIG. 15

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/CA 96/00446

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 H01L29/788 H01L29/10

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 H01L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US,A,5 130 769 (KUO CLINTON C K ET AL) 14 July 1992 see column 3, line 4 - column 3, line 64; figures 6-8	1,3-5
A	EP,A,0 642 172 (PHILIPS ELECTRONICS NV) 8 March 1995 see abstract	8-14

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

- \* "A" document defining the general state of the art which is not considered to be of particular relevance
- \* "E" earlier document but published on or after the international filing date
- \* "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \* "O" document referring to an oral disclosure, use, exhibition or other means
- \* "P" document published prior to the international filing date but later than the priority date claimed

- \* "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \* "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \* "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- \* "&" document member of the same patent family

Date of the actual completion of the international search

20 September 1996

Date of mailing of the international search report

11.10.96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+ 31-70) 340-3016

Authorized officer

Albrecht, C

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/CA 96/00446

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US-A-5130769	14-07-92	JP-A- 5206477	13-08-93
-----			
EP-A-0642172	08-03-95	BE-A- 1007475	11-07-95
		JP-A- 7094613	07-04-95
-----			